Generative Graphical Models for Maneuvering Object Tracking and Dynamics Analysis

Xin Fan$^{1,2}$
$^1$School of Information Engineering
Dalian Maritime University, Dalian, China
xin.fan@okstate.edu

Guoliang Fan$^2$
$^2$School of Electrical and Computer Engineering
Oklahoma State University, Stillwater, OK, USA
guoliang.fan@okstate.edu

Abstract

We study the challenging problem of maneuvering object tracking with unknown dynamics, i.e., forces or torque. We investigate the underlying causes of object kinematics, and propose a generative model approach that encodes the Newtonian dynamics for a rigid body by relating forces and torques with object’s kinematics in a graphical model. This model also accommodates the physical constraints between maneuvering dynamics and object kinematics in a probabilistic form, allowing more accurate and efficient object tracking. Additionally, we develop a sequential Monte Carlo inference algorithm that is embedded with Markov Chain Monte Carlo (MCMC) steps to rejuvenate the path of particles. The proposed algorithm can estimate both maneuvering dynamics and object kinematics simultaneously. The experiments performed on both simulated and real-world data of ground vehicles show the robustness and effectiveness of the proposed graphical model-based approach along with the sampling-based inference algorithm.

1. Introduction

Many applications ranging from intelligent surveillance to military guidance involve the challenging problem of tracking maneuvering objects. One of the major challenges lies in modelling the relationship between maneuvering dynamics and object kinematics. Especially, the former one reflects human’s intentions that are usually unpredictable in a general sense. During the past decades, researchers in computer vision and signal processing communities have made tremendous efforts on dynamic motion modelling, target/background representation and high-dimensional state estimation, etc. It is very important to devise an elaborate motion model when the target is highly maneuvering or there is no accurate appearance model available. A comprehensive survey on motion modelling can be found in[11] where it was stated that “a good model is worth a thousand pieces of data”.

This research starts from the dynamics of a maneuvering object. For many real-world objects, especially man-made ground vehicles and aircrafts, the maneuvering actions are due to the forces and torques in the object’s mechanical driving system[8]. The dynamics of a system demonstrates how the forces and torques cause the kinematic quantities to vary over time, e.g., velocities and positions, which the tracking algorithms try to estimate. The study of the dynamics is based on the physical laws that govern the object motion. These physical constraints allow us to construct a good dynamic/motion model that is capable of accommodating various maneuvering behaviors. Many existing tracking algorithms start from the analysis of object’s dynamics as we do. However, most of them treat the object of interest as a point target [11]. This assumption may not be appropriate any more if the orientations or aspect variations have to be taken into account in the motion model. Instead, in this work, we investigate the full dynamics of a rigid body, in which both linear and angular motions are integrated into one formulation [8].

A graphical model provides an intuitive tool to represent the underlying probabilistic structure of a complex system [10]. We propose a generative graphical model to encode how the forces and torques alter kinematics. This model involves two kinds of latent variables, i.e., the kinematics (effect variables) and the forces/torques (cause variables), and the image sequences are considered as the observations associated with latent states. The conditional densities between the hidden variables are determined by the Newtonian dynamics. Hence the dynamics underlying the maneuvering actions are incorporated into the motion model in a probabilistic way. Moreover, the physical constraints between kinematics and dynamics can be accommodated via conditional probabilities between the latent variables. We use a sequential Monte Carlo (SMC) method to infer the

---

$^1$Kinematics is the study of movement over time, and dynamics is the study of forces that cause the kinematics quantities to change over time.
posterior densities of latent variables given observations, and the Markov Chain Monte Carlo (MCMC) step is used to rejuvenate the distribution of particles. The states of kinematics as well as those of forces/torques reflecting maneuvering actions can be simultaneously obtained.

2. Related work

There exist a large number of motion (dynamic) models proposed for maneuvering object tracking [11], among which the simplest but popular one is the white Gaussian noise acceleration (WGNA) model [4, 6]. As mentioned before, most dynamic models are based on the point target assumption. Miller et al. [13, 12] utilized the Newtonian equations of 2D rigid body dynamics as the motion model and proposed a joint tracking and recognition algorithm. They provided an elegant and unified framework for explicit modeling the object’s movement and rotation upon arbitrary dynamics, and proposed a via the jump-diffusion process for statistical inferencing in the high-dimensional kinematic state space. Since the state space grows exponentially with respect to the length of the observation sequence, the computational complexity is extremely high for practical tracking problems. In order to support sequential inference for state estimation, graphical model-based approaches were invoked where the Markovian structure is assumed and the physical constraints between maneuvering dynamics and object kinematics can also be easily accommodated.

Graphical model-based motion models, e.g., Switching Linear Dynamic (SLD) Systems, have been widely studied for tracking objects with unknown maneuvering behaviors [9, 14, 15]. These models exhibit a similar structure, where a discrete Markov process is employed to generate maneuvering switches among a finite set of continuous linear dynamic models. The probabilistic characteristics of the Markov process are learnt a priori from pre-collected training data. In the same spirit, Doucet et al. [7] proposed the Jump Markov Linear (JML) systems that, however, is not explicitly represented by graphical models. In their models, a discrete Markov process is plugged into the linear dynamic systems as a driving process. One common characteristics of these methods is that the maneuvering dynamics is not explicitly represented and the maneuvering actions are coped with by a discrete switch variable. In this work, we are interested in the direct and explicit dynamic modeling with the aim to deal with arbitrary motion patterns.

Exact inference techniques, e.g., belief propagation [17], are not applicable to the graphical models that have complicated structures, such as the ones describing the motion of maneuvering objects. Researchers resort to the sampling based approximations, e.g., particle filters (SMC) [7] or data-driven Markov Chain Monte Carlo (DDMC) methods [14]. The SMC method can easily propagate particles in a sequential manner according to the motion model for recursive Bayesian state estimation. In addition, the MCMC step can be embedded into the SMC process to overcome the impoverishment problem of the particle set [1]. Hereby we adopt the SMC-based inference algorithm where the physical constraints between maneuvering dynamics and object kinematics can also be incorporated via the MCMC steps.

3. Dynamics on rigid object motion

We start from the dynamics of 3D rigid motion, which reveals the physical constraints of the ground vehicle motion, a special case of 3D motion, as discussed later. The physical laws, also known as Newton equations, which govern the motion of a rigid object are [8]:

\[
\begin{align*}
\dot{p} &= f, \\
\dot{h} &= \tau,
\end{align*}
\]

where \(p\) and \(h\) are the linear momentum and the angular momentum of the body respectively, and \(f\) and \(\tau\) are the force and the torque acted on the object. It should be noted that the Newton equations (1) is valid when the motion is resolved in a fixed inertial reference frame, in which the linear momentum and the angular momentum are dependent on the linear and angular velocities respectively. The dynamics (1) indicates that forces and torques cause the motion of a rigid object whose kinematics is represented by position \(r\), linear velocity \(v\), orientation and angular velocity \(\omega\). We construct our generative graphical model by considering how forces and torques generate kinematical state changes.

In order to characterize the motion of a ground vehicle, we define the three principal axes of the object as the body frame (Fig. 1) and express the dynamics (1) as [12]:

\[
\dot{v} + \omega \times v = \frac{f}{M},
\]

and

\[
J\omega + \bar{\omega} J\omega = \tau,
\]
where \( f = [f_x, f_y, f_z]^T \) and \( \mathbf{v} = [v_x, v_y, v_z]^T \) are composed of the forces and the linear velocities along the body axes, and \( \mathbf{r} = [\tau_x, \tau_y, \tau_z]^T \) and \( \mathbf{\omega} = [\omega_x, \omega_y, \omega_z]^T \) have the torques and the angular velocities w.r.t. the body axes. The mass \( M \) in (2) and the inertial momentum matrix \( \mathbf{J} \) in (3) are fixed for a specified object. And \( \mathbf{\omega} \) is a skew-symmetric matrix with the form of:

\[
\mathbf{\omega} = \begin{bmatrix}
0 & -\omega_z & \omega_y \\
\omega_z & 0 & -\omega_x \\
-\omega_y & \omega_x & 0
\end{bmatrix}.
\]

The main cause of the ground vehicle motion attributes to two variables, \( f_x \) and \( \tau_z \), by analyzing the forces and torques acted on a ground vehicle in the body frame shown in Fig. 1 where a vehicle is abstracted as a cube. We list the detailed analysis on the forces and torques as follows.

- **The driving force,** \( f_z \): The total force parallel to this direction is the summation of engine boosting and friction, i.e., driving or resistence forces. The variations of the force generate the changes of the linear velocity.

- **The sliding force,** \( f_y \): Although there would be a decomposed component of the object gravity when the vehicle moves at a slope, the resistance from the ground would prevent the vehicle from side sliding. Thus, the total force along this direction would be zero.

- **The supporting force,** \( f_x \): It is unlikely for a ground vehicle to jump up and down vertically even when the vehicle is moving at a bumpy terrain. There would be an equilibrium of the forces along this direction, therefore, it is reasonable for us to enforce this force to zero.

- **The rolling torque,** \( \tau_z \): This torque would cause the vehicle’s rolling, which seldom happens for ground vehicles. So we set it zero.

- **The pitching torque,** \( \tau_y \): This torque would cause the vehicle’s pitching, which occurs when the vehicle moving along bumpy pavements. Currently, we assume that the vehicle moves within one plane and enforce this torque as zero.

- **The turning torque,** \( \tau_x \): This torque turns the moving direction. Along with \( f_x \), it reflects the maneuvering actions acting on the object.

Based on the above analysis, equations (2) and (3) can be simplified as:

\[
\begin{bmatrix}
v_x \\
v_y
\end{bmatrix} + \begin{bmatrix}
0 & -\omega_z \\
\omega_z & 0
\end{bmatrix} \begin{bmatrix}
v_x \\
v_y
\end{bmatrix} = \begin{bmatrix}
1 \\
0
\end{bmatrix} f_x \quad (4)
\]

and

\[J_2 \omega_z = \tau_z. \quad (5)\]

![Figure 2. The proposed generative graphical motion model.](image)

Also the the position of the object is related to the velocities \( v \) and \( \mathbf{\omega} \) as:

\[
\dot{\mathbf{r}} = \mathbf{A}(\phi) \mathbf{v} = \begin{bmatrix}
\cos \phi & \sin \phi \\
-\sin \phi & \cos \phi
\end{bmatrix} \begin{bmatrix}
v_x \\
v_y
\end{bmatrix}, \quad (6)
\]

where \( \phi \) is the orientation of the object, and we have

\[
\dot{\phi} = \omega_z. \quad (7)
\]

The differential equations (4), (5), (6) and (7) show how \( f_x \) and \( \tau_z \) drive the kinematical state to change over time. It is worth re-stating that many motion models, including the widely used white-Gaussian-noise acceleration (WGNA) models [6] for multi-aspect target tracking, are obtained from the Newton equations of the particle motion where the object is considered as a point with mass. However, the WGNA models neglect the constraints between linear and angular velocities that can be captured in the dynamics of the rigid body motion model adopted here. Equation (4) shows that the linear acceleration would not occur while the object has non-zero angular velocities. These constrains (causes) can be observed in the motion of ground vehicles so widely that we regard the moving ground vehicles as rigid bodies instead of simple particles.

### 4. Generative graphical models

We develop a generative graphical model to describe the uncertainties of maneuvering actions in Fig. 2 where the object’s dynamics is derived from (4), (5), (6) and (7).

#### 4.1. Probabilistic models of maneuvering variables

Similar to the idea of Jump Markov Linear (JML) models [7] or Switching Linear Dynamic (SLD) models [15], we define multiple models \( M_i \) (\( i = 0, 1, 2 \)) for the driving
force $f_s$, aiming at describing constant velocity, acceleration and deceleration. Each model admits a normal density of $\mu_i$ and $\sigma_i^2$. We update the model parameters during the inference rather than learn them a priori as the one in [13]. Additionally, we specify a prior transition probability matrix (TPM) $T_{ij} = Pr(M_j|M_i)$, e.g., $T_{ij} = 0.5$ when $i = j$ and $T_{ij} = 0.25$ when $i \neq j$.

We assume that the torque $\tau_c$ randomly generates a certain angular velocity $\omega_z$ at a single time step and the angular velocity keeps constant within the sampling interval. Thus, we do not model the torque’s behavior but the distribution of the angular velocity $\omega_z$ by the following ternary-uniform mixture [11]; the object would keep the direction ($\omega_z = 0$) with probability $P_0$; rotate counter clock-wise with probability $\frac{1-P_0}{2}$ at a rate uniformly distributed in $[0, \omega_{max}]$; or clock-wise with probability $\frac{1-P_0}{2}$ at a rate with a uniform distribution in $[-\omega_{max}, 0]$. The value of $P_0$ may reflect some prior knowledge about the possible motion pattern.

4.2. Temporal constraints

The temporal constraints relating the kinematical variables $K$ between two consecutive time steps, where we denote the set of the hidden variables $v, r, \phi$ in Fig. 2 as $K$, are determined by (4), (5), (6). We assume that the sampling interval $T$ is small enough so that we can solve the above differential equations by using Euler integration as:

$$r_t = r_{t-1} + T A (\phi_{t-1}) v_{t-1} + n^r_{t-1}, \quad (8)$$

$$v_t = v_{t-1} + T \omega_{z,t-1} v_{t-1} + n^v_{t-1}, \quad (9)$$

where the velocity vector $v$ comprises the velocities along $x$ and $y$ axes and the skew-symmetric matrix $\omega$ is formed by the angular velocity w.r.t. $z$ as those in (4), and

$$\phi_t = \phi_{t-1} + T \omega_{z,t-1} + n^\phi_{t-1}. \quad (10)$$

The noise variables $n$ are i.i.d. Gaussian, which accommodate numerical errors and uncertainties in the motion process.

The conditional probability of the current kinematical variables given the previous ones $p(K_t|K_{t-1})$, graphically represented by the blue arcs in Fig. 2, can be obtained by (8), (9) and (10) and the given noise statistics. If omitting the arcs directed from $\omega$ to $v$, the proposed graphical model would degenerate to the WGNA model or the constant acceleration dynamic model widely used in the context of target tracking [6, 11]. Besides the temporal evolutions of linear and angular motion independently, our model couples the non-linear angular effects into the linear motion.

4.3. Velocity-force constraints

In the JML [7] and SLD [15] models, an independent discrete Markov chain is introduced as the driving process of continuous states, i.e., kinematics in tracking applications. The kinematics evolution provides cues to infer the underlying driving forces, but does not affect the evolution of the driving process. In our study, we notice that engines of any vehicles have limited power and the output power of an engine equals to the force times its velocity [5]. We direct an arc from $v_t$ to $f_t$ as shown in red in Fig. 2 to encode such constraint, so that object kinematics, say velocity, brings the influence to the forces. Actually, the forces and torques of aircrafts are highly determined by their velocities [8], so our proposed model can be easily generalized to the one applicable to both space aircrafts and other ground vehicles.

We model the conditional probability $p(f_t|v_t, f_{t-1})$, depicted by the red arcs in Fig. 2, as a conditional Rayleigh distribution shown in Fig. 3. The conditional Rayleigh distribution can be regarded as a probabilistic description of how the driving/resistant forces are dependent on kinematics they generates. As seen in Fig. 3, as far as an object being accelerated is concerned, the object is more likely to have higher acceleration (larger driving forces) in a lower speed but prone to moving with lower acceleration if it is moving fast. On the other hand, when an object is in the process of deceleration, the object moving at a higher speed has a high probability to have large resistant forces while a slow-moving object tends to be decelerated gradually with small resistant forces. These considerations are accordant with what we can observe in reality. The distribution is defined as:

$$p(f_t|v_t, f_{t-1}) = \begin{cases} 
\frac{(f_{max} - f_t)}{c_t^2} \exp(-\frac{(f_{max} - f_t)^2}{2c_t^2})1(f_{max} - f_t) & f_{t-1} > 0; \\
\frac{(f_t - f_{max})}{c_t} \exp(-\frac{(f_t - f_{max})^2}{2c_t^2})1(f_t - f_{max}) & f_{t-1} < 0,
\end{cases} \quad (11)$$

where $1(\cdot)$ is the unit step function, $f_{max}$ and $f_{max}$ are the forces that generate maximum acceleration and deceleration respectively, and $c_t$ is a parameter dependent on the current velocity $v$ as:

$$c_t = \begin{cases} 
\frac{f_{max}\sqrt{2/\pi}}{1+\exp(-\sqrt{(2v_{max})^2})} & f_{t-1} > 0; \\
\frac{f_{max}\sqrt{2/\pi}}{1+\exp(-\sqrt{(2v_{max})^2})} & f_{t-1} < 0.
\end{cases}$$

Early in 1984, Zhou and Kumar proposed a mean-adaptive acceleration model as the Rayleigh distribution [18] similar to (11). However, the proposed Rayleigh density is only dependent on previous acceleration instead of current velocity as we do. Additionally, they use this density to derive the variance of the acceleration prediction for Kalman estimation. It can be seen later that we generate samples of the distribution in (11).

Our model shares some common characteristics with the idea of data-driven sampling approaches [16, 14] that the feedback constraints of the effect (e.g., the velocity) on the cause (e.g., the force) is considered. But the data-driven
gives the procedure to update the particle set, in which the prediction paradigm utilizes the cues from the observed data to define a better proposal that has a larger overlap with the target distribution. As shown in Fig. 2, this consideration is explicitly reflected by the conditional distribution between the latent variables in our generative model.

5. Inference algorithm

It is not trivial to obtain the exact inference by standard algorithms, e.g., belief propagation (BP)[17], due to the complicated topology of the proposed model. We resort to a sequential Monte-Carlo (SMC) based algorithm [3] to estimate the kinematics and the underlying probabilistic models of forces and angular velocities.

Noticing that the evolution of the kinematic variables $K$ presents a Markov chain and non-linearity is involved in the dynamic equations, e.g., (8), we can use the SMC algorithm to sequentially maintain the posterior density $p(K_t|Z_{1:t})$ at the current time step given the previous density $p(K_{t-1}|Z_{t-1})$ by a particle set $\{K_i^t, w^t_i\}^N_{i=1}$. Table 1 gives the procedure to update the particle set, in which the prediction $p(K_t^i|K_{t-1}^j)$ is implemented by (8), (9) and (10). And the weights $w^t_i$ are evaluated by the likelihood function, defined to describes how the observed data match the object template $L$:

$$p(Z_t|K^i_t) = C \exp\left(-\frac{(T_K(Z_t) - L)^2}{2\sigma_o^2}\right),$$

where $C$ is a normalization constant, $\sigma_o^2$ is the variance of the observation noise and $T_K$ is a 2D transformation determined by the samples of position $r_i^t$ and orientation $\phi_i^t$.

<table>
<thead>
<tr>
<th>Table 1. The SMC-based inference algorithm.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initializing</strong> sample set ${M_{i0}, f_{i0}, \omega_{i0}, K_{i0}}^N_{i=1}$</td>
</tr>
<tr>
<td><strong>Prediction</strong></td>
</tr>
<tr>
<td><strong>Weighting</strong></td>
</tr>
<tr>
<td><strong>Normalization</strong></td>
</tr>
<tr>
<td><strong>Selection</strong></td>
</tr>
<tr>
<td><strong>MCMC step</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>End</strong></td>
</tr>
</tbody>
</table>

The above SMC algorithm maintains the posterior density $p(K_t|Z_{1:t})$, which generates samples $f_{i-1}^t$ and $\omega_t^i$, we use the Metropolis algorithm [2] to sample the Rayleigh distribution to obtain the samples of $p(f_t|Z_{1:t})$. One sample is generated by one iteration as follows: firstly sample a proposal density $q(\cdot)$, $f^* \sim q(\cdot)$, and then accept $f^*$ as $p(f_t^i)$ with the probability

$$\min\{1, \frac{q(f_{i-1}^t)}{p(f_{i-1}^t)}\},$$

where the proposal $q(\cdot)$ is a Gaussian density and $p(f_t)$ is the conditional Rayleigh distribution given in (11). The sampling approximation of the conditional Rayleigh distributions are illustrated in Fig. 4. The complete procedure of the inference algorithm is shown in Tab 1.

6. Experimental results

Our experiments were conducted on both simulated data and real world video sequences of moving objects. In the simulated data, the positions and orientations were sequentially generated according to the Newton dynamics of a rigid body. A uniform rectangle block corrupted by additive Gaussian noise acts as the object of interest, and it was positioned into an i.i.d. Gaussian background process according to the pre-computed positions and orientations. In a simulated sequence, we controlled the signal-to-noise ratio (SNR) as low as -0.5dB. Also, we captured the movements of a remote-controlled tank from a top-down view as the real-world data, which have 300 frames (24 frames per second) and present highly maneuvering actions. In addition, we post-processed the sequence so as to lower the
SNR to 0dB. In these low SNR scenarios, the tracker would have to rely more largely on dynamic model than appearance model. We used $N_s = 500$ samples in the all following experiments, and we manually initialized the kinematics at the first frame.

Fig. 3 shows the tracking results of our method compared with the jump Markov particle filter (JMPF) [7] on the simulated data sequences. The JMPF algorithm works well in the first 30 frames but starts to loose track afterwards. The algorithm does not follow the force drop (see Fig. 5(c)) that occurs when the speed of the object reaches the maximum. Though a bit delay presents, our proposed method can catch this drop due to the feedback constraints between velocities and forces in our generative model. As time goes on, the JMPF algorithm failed from the 40th frame when an abrupt turn occurs as shown in Fig. 5(d). As the dynamics of a rigid body couples the angular effects into the linear motion (4), our method can appropriately allocate the samples by following this constraint that leads to accurate estimations of kinematics as Fig. 5(a) and (b) show.

Fig. 6(a) shows the tracked trajectory of the object in the real-world sequence by the proposed method, and Fig. 6(b) and (c) illustrate the estimated evolutions of the velocity and orientation over time respectively. The images with the tracked object annotated are presented in Fig. 6. Compared with the tracking algorithm with a simple dynamical model, i.e., constant velocity constant turn (CVCT) [11], and the same likelihood function (12), our method gives robust tracking in such a long sequence even when the dramatic turns occur from time to time. The proposed dynamic model is able to catch up the dramatic target rotations in contrast to the constant turn model that fails after a few frames. In addition, the velocity drops down while the object turns abruptly at the 175th frame. Our generative model based approach can even handle such an irregular motion mode. However, the algorithm with CVCT model drifts much farther away from the true positions as well as orientation. It should be pointed out that our method sometimes takes the shadow as part of the object but it follows the object’s kinematics well as shown in the 150th frame of Fig. 6. This problem could be solved by incorporating a more delicate observation model other than the simple one in (12), which falls out of the scope of this paper.

7. Conclusions and future work

In this paper, we have addressed a challenging tracking problem where the object of interest is highly maneuvering with unknown dynamics. The major contribution is that we have proposed a new generative graphical model to encode the Newtonian dynamics of the moving object in a probabilistic framework. As the consequence, we are able to explicitly and directly build the cause-and-effect relationship between the latent variables that include both dynamics (forces and torques) and kinematics (velocities and orientation). Additionally, the feedback constraint from the velocity to the forces can be easily accommodated via the conditional distribution between the latent variables in the graphical model. We have also developed a SMC-based inference algorithm which is embedded with the MCMC step to improve the sample distribution for recursive Bayesian estimation. We believe that this generative graphical model is an effective approach to cope with maneuvering tracking problems with an explicit dynamics analysis.

In this work, we have demonstrated the performance of our dynamic model through the experiments on tracking an object from the top-down view. Since the generative graphical model is developed based on the motion dynamics in the context of 3D body frame, we could employ this graphical model to track ground vehicles from any perspective view by incorporating a 3D camera model and a 3D object appearance model. Moreover, we want to further explore the potential of the proposed generative graphical model for joint target tracking and recognition in the future.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive comments that improve this paper. This work is supported by the Army Research Office (ARO) under Grant W911NF-04-1-0221.

References

Figure 5. Comparison with JMPF on the simulated data.

Figure 6. Tracking results on a real-word sequence.


Figure 7. Tracking results on 8 representative frames obtained by (a) CVCT model and (b) our approach, where the object with the tracking gate (in the noise-free image) is shown at the bottom-left corner.


